An Analysis of Curvature Sensing for Large Aperture Adaptive Optics Systems

Mark Milman
David Redding
Laura Needels

Jet Propulsion Laboratory
California Institute of Technology
Pasadena, CA 91109

Abstract

This paper presents analysis of the curvature sensing method and the associated wavefront reconstruction problem for adaptive optics applications to telescopes with large aperture. The analysis includes characterizations of the nonlinear, diffraction, and noise effects for curvature sensing. A comparison of reconstruction performance for curvature and slope sensing is also presented. Simulations based on the optical prescription of the Keck telescope are used to verify the analysis.

0. Introduction. Adaptive optics systems are based on sampling the wavefront entering the telescope pupil, and then quickly restoring the aberrated wave to its unaberrated planar state by a corrective optical element. A key component of this process is the measurement and reconstruction of the wave front. Because wavefront measurement devices do not directly measure the phase, but typically some derivative of it, the estimation (or reconstruction) of the phase from these relative measurements is necesary. Many successful AO systems have already been developed using slope measurements of the wavefront obtained by Shack–Hartmann or shearing interferometry. More recently, successful implementations of low order AO systems employing the method of curvature sensing have been made [12]. This paper presents analysis of the curvature sensing method and the associated wavefront reconstruction problem with particular attention towards application to large aperture telescopes.

Methods of curvature sensing using intensity measurements to approximate the Laplacian are described in the papers [6], [1 O], [11], [16], [17]. Here we will analyze some of the properties of the curvature sensing scheme with particular emphasis on the method as devised by Roddier and coworkers [11]. The analysis presented includes nonlinear, diffraction, and noise effects for curvature sensing. We also briefly discuss its viability as a method for estimating the wavefront normal derivative. The ability to perform this latter function is important from the perspective of the sensor operating as a total stand-alone measurement device since the reconstruction of the wavefront from curvature measurements also requires measurements of the normal derivative. Having both the Laplacian and normal derivative on the boundary is required for posing the standard Neumann problem that arises in connection with wavefront reconstruction. A comparison of reconstruction performance for curvature and slope sensing is presented.

Briefly, the method of curvature sensing relics on forming a normalized difference of intensities in two " planes symmetrically displaced from the focal plane. Using a Fresnel propagation analysis, and with several approximations, including "seeing" condition assumptions, the irradiance transport equation is derived. The irradiance transport equation is essentially a hydrodynamic model describing the evolution of the intensity of the wavefront in the direction of propagation. The derivative of the intensity in the direction of propagation is shown to be proportional to the Laplacian of the wavefront surface. Thus intensity measurements in displaced planes normal to the propagation provide a finite difference approximation to the differential intensity, and hence to the wavefront Laplacian.

Within a geometric model of intensity propagation [2] it is shown that the nonlinearities of the sensor can be characterized through the Gaussian curvature of the wavefront [8]. The scale of this nonlinearity grows with the sensitivity of the measurement. An expression characterizing the balance between the sensor nonlinearity and noise characteristics is derived. It is shown that diffraction effects can be incorporated via a convolution of the curvature signal with certain impulse response functions of the telescope. For large aperture systems these effects are relative] y benign, although simulation and analysis indicate that they do have impact in the neighborhood of obscurations and and at the pupil edges. The analysis of radial derivative measurements obtained by intensity measurements indicates that it is susceptible to nonlinearities and noise effects. This measurement is concluded to be somewhat inferior to a Hartmann sensor measurement.

Reconstruction error covariance matrices are developed for both slope sensing and curvature sensing. Covariance matrices of the reconstruction error for both sensing methods are developed. For the case of square arrays analytic expressions are given for the reconstructed wavefront variance. It is shown that these variances are determined from the eigenvalues of the Laplacian operator discretized by a 5–point scheme over the array. Because these eigenvalues are known for the square geometry, the variances can be simply calculated and comparisons can be made between curvature and slope sensing performance. The results comparing these sensing methods are somewhat contrary to those previously reported [11], [13], [14]. It was observed in numerical studies that although the Hartmann sensor estimation error was superior to curvature sensor estimation error, curvature sensing compared more favorably as the number of sensors *increased* as opposed to decreased, as has been reported in these references. This phenomenon is explained in terms of the *accuracy* (in the sense that the Laplacian is a second order measurement while the gradient is a first order measurement) of the two sensing methods, and an asymptotic expression is given for the estimation variance for both schemes.

1. Nonlinear Geometric Analysis. Consider the diagram below. Here P represents a fixed plane transverse to the direction of propagation, and $P_{\pm}$ are parallel planes displaced a distance $\pm\epsilon$ from P. The function w represents the wave front error function, the deviation of the wave front from being a plane wave .

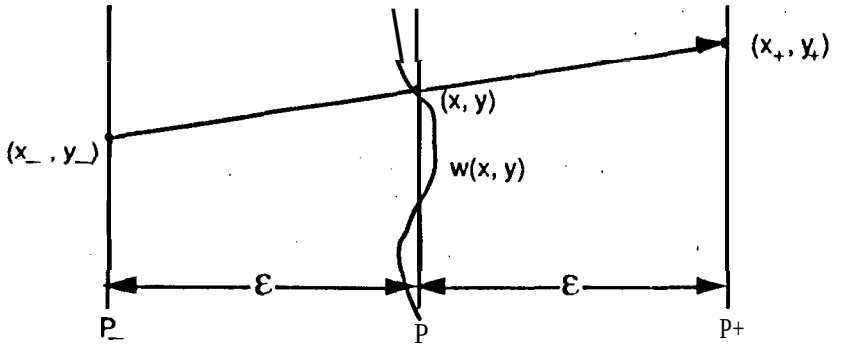

Figure 1.1. Geometric Intensity Propagation

propagating in the z– direction. Let $I_0(x, y)$ denote the intensity of the signal at a point $(x, y) \in P$. From geometric optics, the intensity $I_\pm$ at a point $(x_\pm, y_\pm) \in Pi$ displaced a distance $\pm\epsilon$ along the normal to the wavefront surface is given by

$$I_\pm = \frac{I_0}{1 \mp 2\epsilon H + \epsilon^2 K}, \tag{1.1}$$

where H and K are the mean and Gaussian curvatures of the surface, respectively,

$$H = \frac{\Delta w}{\sqrt{1 + w_x^2 + w_y^2}}, \quad K = \frac{w_{xx} w_{yy} - [w_{xy}]^2}{\sqrt{1 + w_x^2 + w_y^2}}. \tag{1.2}$$

Here Δw denotes the Laplacian, $\Delta w = w_{xx} + w_{yy}$. Expression (1.1) can be deduced from the geometric intensity propagation results in [2] together with the classical formulas for mean and Gaussian curvature given in [8]. Define the normalized difference Q as

$$Q = \frac{I_+ - I_-}{I_+ + I_-} \tag{1.3}$$

Now, $I_+ - I_-$ is easily computed to be

$$\begin{aligned} I_+ - I_- &= I_0\left\{\frac{1}{1 - 2\epsilon H + \epsilon^2 K} - \frac{1}{1 + 2\epsilon H + \epsilon^2 K}\right\} \\ &= \frac{4 I_0 \epsilon H}{1 - 4\epsilon^2 H^2 + 2\epsilon^2 K + \epsilon^4 K^2}. \end{aligned} \tag{1.4}$$

Similarly we find that

$$I_+ + I_- = \frac{2 I_0 (1 + \epsilon^2 K)}{1 - 4\epsilon^2 H^2 + 2\epsilon^2 K + \epsilon^4 K^2}. \tag{1.5}$$

Hence,

$$Q = \frac{2\epsilon H}{1 + \epsilon^2 K} \tag{1.6}$$

Equation (1.1) holds so long as $1 \mp 2\epsilon H + \epsilon^2 K < 0$. (This condition essentially precludes caustics.) Thus, the expression for Q above is valid under the same condition. Note that as $\epsilon \to$ O, Q/c'4 2H. And since $I_+ - I_- = 2I_0 + O(\epsilon^2)$, it follows that

$$\lim_{\epsilon \to 0} \frac{I_+ - I_-}{2\epsilon} \frac{1}{I_0} = 2H. \tag{1.7}$$

But

$$\lim_{\epsilon \to 0} \frac{I_+ - I_-}{2\epsilon} = \text{VI. } n,$$

where

$$\frac{[-w_x, -w_y, 1]^T}{\sqrt{1 + w_x^2 + w_y^2}}$$

is the unit normal to the wavefront surface. Thus after imposing the paraxial assumption

$$|\frac{\partial w}{\partial x}|, |\frac{\partial w}{\partial y}| << 1,$$

wc obtain

$$\frac{\partial I}{\partial z} = \text{VI}, \nabla w + 2I_0 H.$$

And since the paraxial assumption also implies $H \approx \Delta w/2$, the transport equation

$$\frac{\partial I}{\partial z} = \text{VI} \cdot \nabla w + I_0 \Delta w. \qquad (1.8)$$

is obtained. The transport equation above is also valid under paraxial physical optics assumptions [17].

The sensor signal Q is typically modeled not by (1.6) but by the transport equation (1.8)} and the validity of the model requires a "small" displacement between the planes where the intensity measurements arc taken. It will be seen that instrument sensitivity requires a "large" displacement between these planes. This has two effects. First, the nonlinearities become more significant and the transport model (1.8) looses validity. And second, the role of diffraction becomes more dominant in the actual instrumentation setup. Here wc will usc (1.6) as the departure point for the subsequent analysis. The analysis will begin from a geometric optics perspective with the nonlinearities described above, and then noise and diffraction effects will bc included. We will proceed here by analyzing the Roddier set–up [10]. The model for this approach is shown below in Figure 3.2.

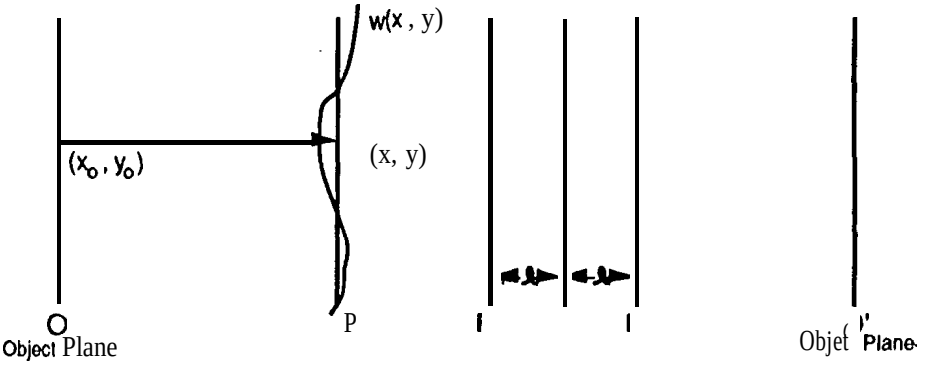

Figure 1.2. Geometric Intensity Propogation in Instrument

Here F denotes the focal plane, F_+ is the image plane associated with the object plane O, and F- is an "image" plane associated with an object plane O'. The planes F_- and F_+ are both a distance 1 from the focal plane F. Let (x, y) be a point in the pupil plane P and let n denote the normal derivative to the wavefront in the direction of propagation,

$$n = \frac{[-w_x, -w_y, 1]^T}{\sqrt{1 + w_x^2 + w_y^2}}$$

Let (x_0, y_0) denote the point in the plane O that contains the light ray through (x, y). The coordinates of (x_0, y_0) arc

$$(x_0, y_0) = (x + \frac{f(f+l)}{l} w_x, y + \frac{f(f+l)}{l} w_y).$$

Let ρ_+ denote the distance between the points $(cc_o, y_0) \in O$ and $(z, y) \in P$:

$$\rho_+ = \frac{f(f+l)}{l}\sqrt{1 + w_x^2 + w_y^2}.$$

Then using geometric optics [Born], the intensity at (x_0, y_0) is

$$I(x_0, y_0) = \frac{I(x, y)}{1 + 2\rho_+ H_w + \rho_+^2 K_w}, \tag{1.9}$$

where H_w and K_w are the mean and Gaussian curvatures of the wavefront, respectively, and $I(x, y)$ is the intensity at (z, y). Next let (z_+, y_+) denote the Gaussian image of (x_0, y_0). Then

$$(x_+, y_+) = (-lx/f - (f+l)w_x, -ly/f - (f+l)w_y). \tag{1.10}$$

Also, the intensity at (z_+, y_+) is given by

$$I_+(x_+, y_+) = (\frac{f}{l})^2 I(x_0, y_0)$$
$$= \frac{I(X, y)}{(+)\,2\,[]+ \frac{f(f+l)}{l}\Delta w - t \frac{f^2(f+l)^2}{l^2}(w_{xx}w_{yy} - [w_{xy}]^2)\}} \tag{1.11}$$

The expression above describes the intensity in the image plane F_+ predicted by geometric optics. Next we will derive the intensity in the plane $F-$ under the same approximations.

Note that the wavefront emerging from the entrance pupil is given by

$$\tilde{w}(x, y) = w(x, y) + \frac{x^2 - t - y^2}{2j}. \tag{1.12}$$

Tracing the ray through (z, y) to the plane F_, we find that it pierces this plane at $(z-, y_-)$,

$$(x_-, y_-) \cdot (x - (j - l)\tilde{w}_x, y - (f - - l)\tilde{w}_y)$$
$$= (lx/f - (f \cdot - l)w_x, ly/f - (j - l)w_y). \tag{1.13}$$

Consequently, the intensity at $(z.., y_-)$ is given by.

$$I_-(x_-, y_-) = \frac{I(Z, y)}{1 - 2p_- H_{\tilde{w}} + \rho_-^2 K_{\tilde{w}}}, \tag{1.14}$$

where $p- = (j - l)\sqrt{1 + \tilde{w}_x^2 + \tilde{w}_y^2}$, and $H_{\tilde{w}}$ and $K_{\tilde{w}}$ are the mean and Gaussian curvatures of the wavefront $\tilde{w}$. But now note that

$$2H_w = \frac{\Delta w + 2/f}{\sqrt{1 + \tilde{w}_x^2 + \tilde{w}_y^2}},$$

$$K_{\tilde{w}} = \frac{1}{1 + \tilde{w}_x^2 + \tilde{w}_y^2}\{\tilde{w}_{xx}\tilde{w}_{yy} - [w_{xy}]^2\}$$
$$= \frac{1}{1 + \tilde{w}_x^2 + \tilde{w}_y^2}\{(w_{xx} + 1/f)(w_{yy} + 1/f) [w_{xy}]^2\}$$
$$= \frac{1}{1 + \tilde{w}_x^2 + \tilde{w}_y^2}\{[w_{xx}w_{yy} - [w_{xy}]^2] + 1/f^2 + \Delta w/f\}.$$

Hence,

$$I_-(x_-,y_-) = \frac{I(x,y)}{1-(f-l)(\Delta w+2/f)+(f-l)^2(1/f^2+\Delta w/f+w_{xx}w_{yy}-[w_{xy}]^2)}$$
$$= \frac{I(x,y)}{(1-(f-l)/f)^2+(f-l)\Delta w\{(f-l)/f-1\}+(f-l)^2\{w_{xx}w_{yy}-[w_{xy}]^2\}} \quad (1.15)$$
$$= \frac{I(x,y)}{(\frac{l}{f})^2\{1-f(f-l)\Delta w/l+f^2(f-l)^2/l^2\{w_{xx}w_{yy}-[w_{xy}]^2\}}.$$

Thus we may compute the normalized intensity difference as

$$\frac{I_+(x_+,y_+)-I_-(x_-,y_-)}{I_+(x_+,y_+)+I_-(x_-,y_-)} = \frac{\frac{f^2}{l}\{-2\Delta w+f[w_{xx}w_{yy}-[w_{xy}]^2]\}}{2+2f\Delta w+\frac{f^2}{l^2}[(f+l)^2+(f-l)^2][w_{xx}w_{yy}-[w_{xy}]^2]}.$$

Let $t = f^2/l$, and assume 1<< j. Then the expression above can be written as

$$\frac{I_+(x_+,y_+)-I_-(x_-,y_-)}{I_+(x_+,y_+)+I_-(x_-,y_-)} = \frac{-t\Delta w+lt^2[w_{xx}w_{yy}-[w_{xy}]^2]/f}{1+lt\Delta w/j+t^2[w_{xx}w_{yy}-[w_{xy}]^2]}$$

For the scale and type of aberrations of interest for the Keck telescope application we will typically have

$$|t^2[w_{xx}w_{yy}-[w_{xy}]^2]| \approx |t\Delta w|; \quad //j << 10^{-2}.$$

Hence wc have approximately

$$\frac{I_+(x_+,y_+)-I_-(x_-,y_-)}{I_+(x_+,y_+)+I_-(x_-,y_-)} \approx \frac{-t\Delta w}{1+t^2K_w} \quad (1.16)$$

Note first that even if intensity measurements are taken along the normal to the wavefront, a nonlinearity involving the Gaussian curvature K_w of the wavefront is encountered. Furthermore, unless Vw = O, an additional nonlinearity emerges since $(x_-, y_-) \neq (x_+, y_+)$. However, this latter error can be shown to be negligible [7].

2. Noise Effects, Assume a sensor integration time of AT seconds, and let A denote the area of the detector element, in the planes F_- and F_+. The number of photons captured by these detectors over the period AT is modeled as independent Poisson processes N_+ and N_- with means $\bar{N}_+$ and $\bar{N}_-$, respectively. The mean signal intensities in the two planes are then

$$I_\pm = \frac{\bar{N}_\pm}{ATA}.$$

Let $I_0 = (I_+ + I_-)/2$, and define the random variable $\hat{S}$ by

$$\hat{S} = \frac{N_+ - N_-}{2I_0\Delta TA}, \quad (2.1)$$

Then S is an estimate of the normalized intensity difference with mean

$$E[\hat{S}] = \frac{I_+ - I_-}{I_+ + I_-}, \quad (2.2)$$

and variance (using $N_\pm$ are Poisson)

$$E[(\hat{S} - E(\hat{S}))^2] = E(\{\frac{N_+ - N_- - (\bar{N}_+ - \bar{N}_-)}{\bar{N}_+ + \bar{N}_-}\}^2)$$
$$= \frac{1}{(\bar{N}_+ + \bar{N}_-)^2}E\{(N_+ - \bar{N}_+)^2 + (N_- - \bar{N}_-)^2 + 2(N_+ - \bar{N}_+)(N_- - \bar{N}_-)\} \quad (2.3)$$
$$= \frac{1}{(\bar{N}_+ + \bar{N}_-)^2}[N_+ + \quad]$$
$$= \frac{1}{\bar{N}_+ + \bar{N}_-}.$$

Write ϕ for the nonlinear terms in (1,16). Observe that

$$E(\hat{S}) = t\Delta w + \phi; \phi \approx \frac{t^3 \Delta w K_w}{1 + t^2 K_w}$$

and consequent] y

$$\Delta w = E(\hat{S}/t) + \phi/t.$$

Now define the random variable $\eta = \hat{S}/t - E(\hat{S}/t)$. The measured signal derived from (2.1) is

$$y = \hat{S}/t.$$

Thus wc can write

$$\begin{aligned} y &= E(\hat{S}/t) + [\hat{S}/t - E(\hat{S}/t)] \\ &= \Delta w + \phi/t + \eta, \\ &\approx \Delta w - \frac{t^2 \Delta w K_w}{1 + t^2 K_w} + \eta \end{aligned} \tag{2.4}$$

with $E(\eta) = O$ and $E(\eta^2) = 1/t^2(\bar{N}_+ + \bar{N}_-)$. Hence the balance that must be maintained is to keep the noise level small by choosing t to be as large as possible, while keeping the nonlinearities at bay with t sufficiently small.

3. Diffraction Effects. Now we will move on to' how diffraction affects the curvature measurement. Recall that the impulse response $h_+(z)$ between the object plane O and image plane F_+ is computed as the response at x in the image coordinates to a point source located at the origin in the object plane [4]. The point source gives rise to a diverging spherical wave at the entrance pupil. Diffraction effects in the planes $F_\pm$ will be modeled in this fashion.

Again let $I(x, y)$ denote the intensity in the entrance pupil. The complex amplitude U^+ in the object plane Oat the point (x_0, y_0) is given as (cf (1.9))

$$U^+(x_0, y_0) = \sqrt{I(x_0, y_0)} e^{\phi_+(x_0, y_0)}, \tag{3.1}$$

where $I(x_0, y_0)$ is given by (1.9) and ϕ_+ is the phase calculated from the phase at (x, y) using the distance function ρ_+. With the assumption that the intensity is constant in the entrance pupil, $I(x, y) = I_0$, we get upon retaining the first order terms in the binomial expansion of (1 .9)

$$U^+(z, y_0) = \sqrt{I_0}\{1 - \rho_+ H - \rho_+^2 K/2\} e^{\phi_+}. \tag{3.2}$$

"The amplitude in F_+ is obtained by convolving the complex amplitude obtained by the geometric optics prediction $U_+(x, y) = \frac{f}{l} U^+(-fx/l, fy/l)$, with the impulse response to obtain

$$U_{F_+}(x, y) = \int_{\Sigma_+} h_+(x - z', y - y') U_+(x', y') dx' dy', \tag{3.3}$$

where Σ_+ is the support of U_+. (This is the set where U_+ is nonzero. In this way we will be able to account for obscurations.) The approximation made here is the use of geometric optics propagation between the planes O and P.

Diffraction effects in the plane F. can be computed in a similar manner. To see this observe that the impulse response h_+ can be interpreted as the response to a diverging spherical wave in the entrance pupil with a radius of curvature given by the distance d_0 from the point source to the pupil. Suppose instead we compute the response of a converging spherical wave in the entrance pupil. This converging spherical wave can be thought of as being produced by a point source in the virtual object plane O' located a distance $d_{0'}$ from the pupil plane satisfying the lens condition

$$\frac{1}{(f - l)} + \frac{1}{d_{0'}} - \frac{1}{f} = 0. \tag{3.4}$$

This interpretation gives rise to the impulse response h- computed in exactly the same manner as h_+ *(see [3, 8]),* with the exception that $d_{0'}$ is a ***negative*** quantity because 1>0 above.

Using the same approximations in computing the diffraction effects in the plane F_+, the complex amplitude in F_- is then computed by geometrically propagating the wavefront $w(x, y)$ to the object plane O', and convolving with the impulse response h_-,

$$U_{F_-}(x, y) = \int_{\Sigma_-} h_-(x - z', y - y')U_-(x, y)dx'dy', \tag{3.5}$$

where U- (z, y) is the amplitude predicted by geometric optics

$$U_-(Z, y) = \frac{f}{l}U^-(-fx/l, fy/l) \quad \text{and} \quad U^-(x_0, y_0) = \sqrt{I_0}\{1 - t\ \rho_- H - \rho_-^2 K/2\}e^{\phi_-},$$

and Σ_- is the support of U_-.

Writing the integrals in (3.3) and (3.5) as the convolutions $h_+ * U_+$ and $h_- * U_-$, respectively, the intensities in $F_\pm$ are then $|h_\pm * U_\pm|^2$, and the curvature signal Q, including the effects of diffraction, is given as

$$Q = \frac{|h_+ * U_+|^2 - |h_- * U_-|^2}{|h_+ * U_+|^2 + |h_- * U_-|^2} \tag{3,6}$$

Before carrying these computations forward, we will make some simplifying assumptions and introduce some useful notation. First, with the assumption $1 << f$, we may assume $h_+ \approx h_-$ and write $h_+ = h_- = h$. This assumption also implies that $\rho_+ \approx \rho_-$, and we will accordingly write $\rho_+ = \rho_- = t = l/2$. Another assumption that wc shall make is that the support of the impulse response h is much smaller than the quantity Tel/j where r. denotes the atmosphere coherence length. Taking D to denote the telescope diameter, and λ as the operating wavelength, this assumption is equivalent to requiring

$$\frac{\lambda f}{Dr_0} << l. \tag{3.7}$$

As shown below, this assumption allows us to ignore the phase terms $\phi_\pm$ in the computations of the intensities. First wc introduce the functions a, b defined as

$$a = \frac{f}{l}\sqrt{I_0}[1 - t^2K], \quad b = \frac{f}{l}\sqrt{I_0}tH, \tag{3.8}$$

so that

$$U_\pm = [a \mp b]e^{\phi_\pm}. \tag{3.9}$$

The intensity at a point $u \in F_+$ can now be expressed as

$$\begin{aligned} |(h * U_+)(u)|^2 &= \int_{F_+} h(u - v)[a(v) - b(v)]e^{\phi_+(v)}dv \int_{F_+} \bar{h}(u - v')[\bar{a}(v') - \bar{b}(v')]e^{-\phi_+(v')}dv' \\ &\approx e^{\phi_+}(v)e^{-\phi_+}(v) \int_{F_+} h(u - v)[a(v) - b(v)]dv \int_{F_+} \bar{h}(u - v')[\bar{a}(v') - \bar{b}(v')]dv', \end{aligned} \tag{3.10}$$

where the overbar denotes complex conjugation. The phase terms are seen to cancel and the intensity at u is independent of the phase. The same result holds for $|h * U_-|^2$.

If in addition the pupil function is radially symmetric, h becomes real, and we can ignore taking complex conjugates above. We will make this assumption in what follows. (This assumption simplifies the notation, but has little effect on the analysis.)

Now we introduce the sets

$$\Sigma_1 = \Sigma_+ \cap \Sigma_-, \quad \Sigma_2 = \Sigma_+ - \Sigma_-, \quad \Sigma_3 = \Sigma_- - \Sigma_+. \tag{3.11}$$

Let χ denote the characteristic function of a set, i.e., x(S)(Z) = 1 if $x \in$ S, and zero otherwise. For i = 1,2,3 we let $a_i(u)$ = u p, a n d $b_i(u) = b(u)\chi(\Sigma_i)(u)$. With this notation and the various approximations made, (3.6) becomes

$$Q = \frac{N}{D}, \tag{3.12}$$

where

$$N = 4(h * a_1)(h * b_1) + 2[h * (a_1 + b_1)][h * (a_2 + b_2)] + [h * (a_2 + b_2)]^2 - 2[h * (a_1 - b_1)][h * (a_3 - b_3)] - [h * (a_3 - b_3)]^2,$$

and

$$D = 2[h * a_1]^2 + 2[h * b_1]^2 + [h * (a_2 + b_2)]^2 + [h * (a_3 - b_3)]^2 + 2[h * (a_1 + b_1)][h * (a_2 + b_2)] + 2[h * (a_1 - b_1)][h * (a_3 - b_3)].$$

For points far removed from the pupil edges or obscurations (i.e, well within the interior of Σ_1), we can ignore the contribution to the signal made by the sets Σ_2 and Σ_3 and obtain

$$Q \approx \frac{2(h * a_1)(h * b_1)}{(h * a_1)^2 + (h * b_1)^2}. \tag{3.13}$$

Recalling the definitions of a_1 and b_i(cf (3.8) and (3.11))), and ignoring higher order terms,

$$\begin{aligned} Q &\approx \frac{2t(h * 1)(h * \mathrm{H})}{(h * 1)^2 + (h * 1)(h * t^2K)} \\ &= \frac{th * \Delta w}{(h * 1) + (h * t^2K)}. \end{aligned} \tag{3.14}$$

Finally, introducing the notation

$$\tilde{h} = \frac{h}{\int h}, \tag{3.15}$$

so that $\tilde{h} * 1 = 1$, we obtain the result (for points removed from edges and obscuration)

$$Q \approx \frac{t\tilde{h} * \mathrm{Au})}{1 + t^2\tilde{h} * K}. \tag{3.16}$$

For these points it is seen that the curvature signal is characterized by convolving the individual terms of the geometric model with the point spread function of the instrument. For systems with large aperture, $\tilde{h}$ is an approximate δ function, and wc recover the geometric model.

Modifying the sensor model to include diffraction effects is straightforward. For points in the interior of Σ_1, the estimator is not estimating the Laplacian of the wavefront, Aw, but the convolution of the Laplacian with the normalized point spread function of the instrument. Thus the model becomes

$$y = \tilde{h} * \Delta w + \tilde{h} * \psi + \eta, \tag{3.17}$$

where * again denotes convolution, and ψ and η are the nonlinear and noise terms from (2.4). Closer to the boundaries of the obscurations and pupil edges the full quotient model (3.12) must be used.

An Example. Thus far we have developed how intensity measurements lead to estimates of wavefront curvature and radial tilt. We will now consider some examples using low order Zernike aberrations.

The Zernike polynomial for tilt is

$$w(x, y) = \frac{x}{R} \quad (\text{or } z = \frac{\rho}{R}\cos\theta \text{ in polar coordinates}); \quad \text{R=pupil radius.}$$

Hence,

$$\mathrm{Aw} = 0,$$

and

$$< \nabla w, n >= \frac{cos\theta}{R}.$$

The zero Laplacian of tilt is captured by the intensity signal in Figures 3.1a–3.1 c. In each of these figures we chose the displacement 1 from focus to be .05m, and the focal length of the system as f=150m. Thus t = 4.5x 10^5m. Figure 3.1a contains the Keck prescription without central obscuration from the secondary mirror. The magnitude of the signal increases to unity at the edge of the pupil. Figures 3.1 b-3.1 c contain the signal with obscuration. There is more ringing to the signal in these cases because of the diffraction contribution of the secondary mirror. The signal in the center for Figures 3.1b–3.1c is due entirely to diffraction. Because the aberration consists of an $x-$ axis tilt, the terms containing the Laplacian and Gaussian curvature in (3.12) disappear. Along the $y-$ axis we would expect the signal to diminish in the obscured region because the terms in the numerator cancel. This is precisely the case as can be observed in Figure 3.1c.

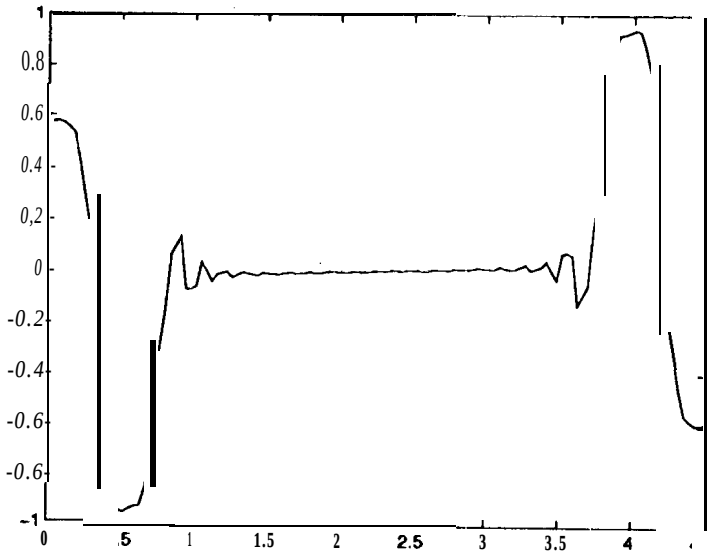

Figure **3.1a.** $x-$ tilt **Laplacian** (no obscuration), $x-$ axis slice

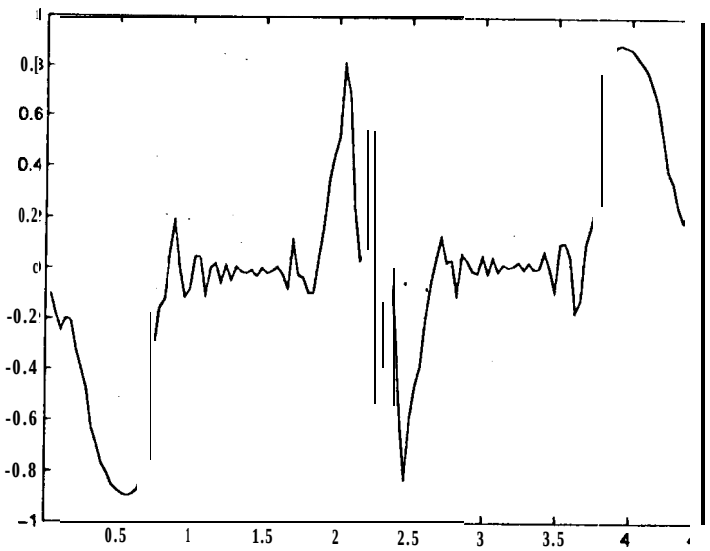

Figure **3.1b.** $x-$ tilt **Laplacian** (with secondary obscuration), $x-$ axis **slice**

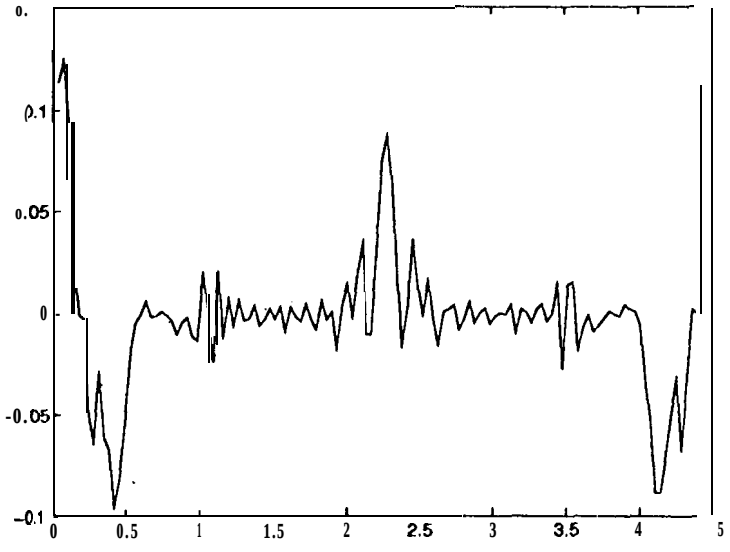

Figure 3. lc. $x-$ tilt **Laplacian** (with secondary obscuration), $y-$ axis **slice**

The Zernike polynomial for defocus is given by

$$W(Z, y) = 2(=)^2 + 2(\frac{y}{R})^2 - 1; \text{ R=pupil radius}$$

Thus with a coefficient of ϵ multiplying the defocus term, the resulting Laplacian and radial tilt are

$$\epsilon\Delta w = \frac{8\epsilon}{R^2},$$

and

$$\epsilon < \nabla w, n >= \frac{4\epsilon}{R},$$

respectivel y.

The geometric curvature signal for defocus is given by

$$\frac{I_+(x, y) - I_-(x, y)}{I_+(x, y) + I_-(x, y)} = \frac{-f_l^2[\frac{8\epsilon}{R^2}]}{1 + (\frac{f_l^2}{l})^2 \frac{16\epsilon^2}{R^4}}.$$

The nonlinearity introduced by the Gaussian curvature term in the denominator is seen in the simulations by comparing intensities from the normalized intensity maps as the image planes are moved closer to focus (Figs. 3.2a-3 .2c). The intensities increase sublinearly because of this term. Table 3.1 summarizes the value of the curvature signal for these cases using the linear model (1.8), nonlinear model (1.6), and simulated intensities (including diffraction effects.) It is seen that as 1 decreases the linear model becomes quite inaccurate, while the nonlinear model retains good fidelity with the simulated intensities.

Table 3.1 **Sublinearity** of curvature **measurements**

l	25mm	**50mm**	**100mm**
Linear	.29	.58	1,16
Geometric	.28	. 5 3	.86
Simulation	.28	**.54**	.87

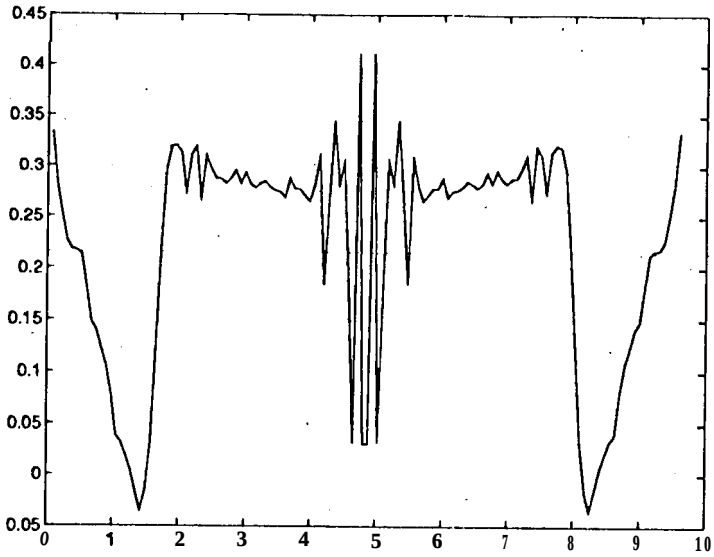

Figure 3.2a. Defocus **Laplacian** *l* = .1 mm

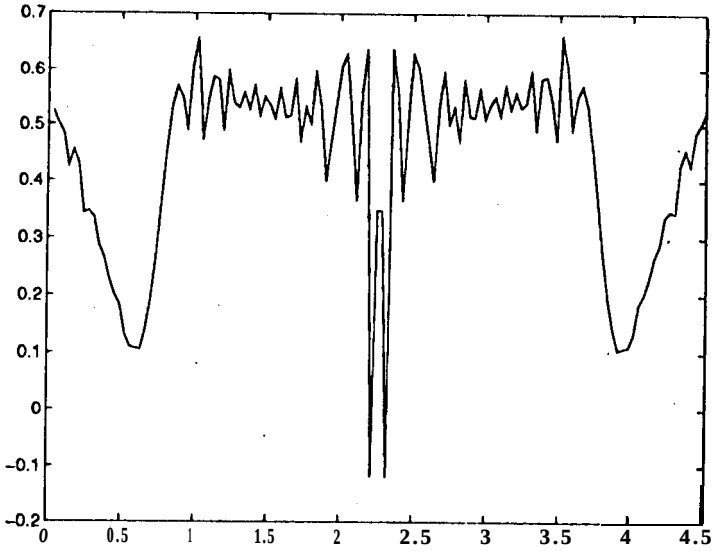

Figure **3.2b.** Defocus **Laplacian** 1 = .05 mm

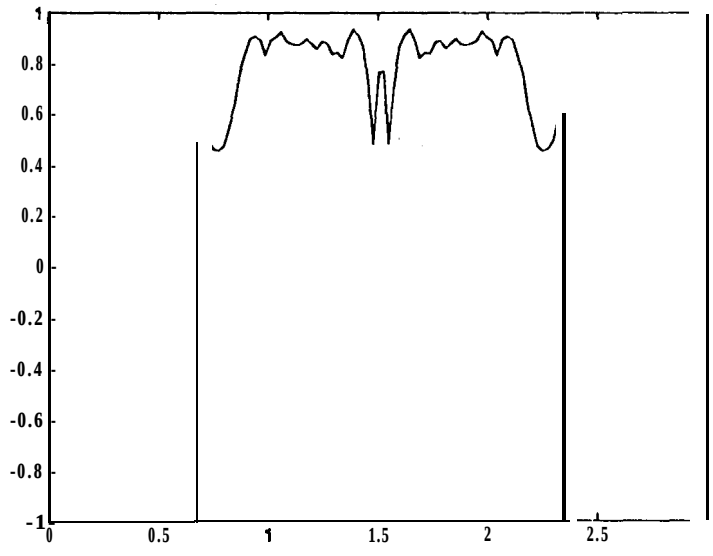

Figure **3.2c.** Defocus **Laplacian** 1 = .025 mm

4. Radial **Derivat** ive from Intensity Measurements. The problem of **wavefront reconstruction** from intensity **measurements** requires a measurement of the boundary slope. Here wc will briefly describe how **the** intensity distribution can provide information regarding the normal derivative of the wavefront on the boundary. This will be done from a geometrical optics perspective, however diffraction effects can be accommodated via (3.1 2). (A more detailed analysis of boundary slope determination can **be** found in [7].)

Let F' denote the pupil, and $\bar{P}$ denote the pupil including its boundary, ∂P. Given a point (z, y) $\in \bar{P}$, with wavefront (x, y, w (x, y)), rays are propagated to the object planes O and O' by the transformations

$$T_{\pm}(x,y) :\rightarrow (x,y) \pm t(\frac{\partial w}{\partial x}, \frac{\partial w}{\partial y}).$$

Introducing unit vectors in the radial and tangential directions, u_r and u_τ, respectively, the transformations above are written as

$$T_{\pm}(x,y) :\rightarrow (x,y) \pm t < \nabla w, u_r > u_r + < \frac{1}{R}\nabla w, u_\tau > u_\tau,$$

where R denotes the aperture radius. From this we see that differentiall y the transformations $T_{\pm}$ consist of a rotation, via u_τ, and a radial translation, via u_r. To compute the effect on the boundaries $T_{\pm}(\partial P)$ *fix* θ_0 and define' '

$$\theta_{\pm} = \theta_0 \mp \text{i?} < \nabla w(R,\theta_0), \text{W} >$$

Now observe that to first order

$$T_{\pm}(R,\theta_{\pm}) = (R_{\pm},\theta_0),$$

where

$$R_{\pm} = \pm t < \nabla w(\theta_{\pm}), u_R(\theta_{\pm}) >$$

So to first order

$$R_{+} - R_{-} = 2t < \nabla w(\theta_0), u_R(\theta_0) > .$$

In the image planes, F_{+} and F_{-}, the radii are magnified by the factor l/j so that

$$l/f(R_{+} - R_{-}) = 2f < \nabla w(R,\theta_0), \text{n} > . \tag{4.1}$$

To summarize, let $S_{\pm} = 1/fT_{\pm}$. Then the radial displacement of the boundaries of $S_{+}(\partial P)$ and $S_{-}(\partial P)$ at a point (R, θ) is proportional to the wavefront slope at the point; and further the proportionality constant .

is the displacement between the planes P_+ and P_-. And in the image planes $F_\pm$ this proportionality constant becomes $2tl/f = 2f$. Note that the displacement depends only on the focal length and not on the position of the image planes.

The determination of radial slope by intensity measurements of the pupil displacements at the boundary is not a direct process. Next we will discuss how this is done in the Roddier scheme.

The Roddier scheme for estimating the radial tilt < Vw, n > is based on the observation that the curvature signal $(I_+ - I_-)/(I_+ + I_-)$ is ± 1 for points in the symmetric difference of the sets $S_\pm(P)$: Points in $(S_+(P) - S_-(P))$ U $(S_-(P) - S_+(P))$, i.e. where there is no overlap, and is strictly less than one where there is overlap. The distance where there is no overlap is proportional to the wavefront tilt on the boundary as we saw above. In Fig. 3,3 the region G is written as the union of two subregions, GO and G_1. Go is bounded by small arcs of $S_+(\partial P)$ and $S_-(\partial P)$, and thus the width of Go is proportional to < Vw, n >.' Since it is not *a priori* known what the width actually is, a parameter μ is chosen as an upper bound. μ is the width of the region G. < Vw, n > is estimated by averaging the observed curvature signal over the entire region G. Obtaining an accurate estimate of < Vw, n > with this procedure relies on a small value of the (curvature) signal in G_1. (This can be enforced by choosing large 1; however this increases the size of the imaged pupil, and thereby reduces the sensitivity of the tilt measurement since the relative displacement of the pupil boundary is independent of l.)

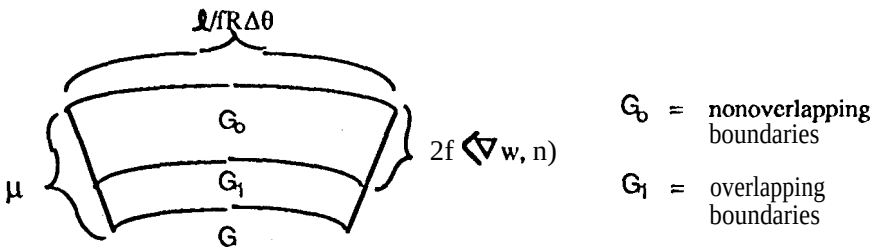

Figure 4.1. Normal Derivative Determination

Now let $\widehat{\partial w}$ denote the estimate of < Vw, n > obtained by averaging the normalized intensity signal over G:

$$\widehat{\partial w} = \frac{1}{2\,f\frac{l}{f}R\Delta\theta}\int_G \frac{I_+ - I_-}{I_+ + I_-} = \frac{1}{2lR\Delta\theta}\int_G \frac{I_+ - I_-}{I_+ + I_-}$$

Partitioning G into the regions indicated in the figure above, G = Go $\bigcup$ G_1, and performing the integrations, wc obtain

$$\int_G \frac{I_+ - I_-}{I_+ + I_-} \approx 2fR\Delta\theta l\frac{<\nabla w, n>}{f} + \delta\alpha R(l/f)(f^2/l)\Delta\theta\Delta w,$$

where $\delta\alpha = \alpha - 2f <\nabla w, n>$: Hence,

$$\widehat{\partial w} \approx <\nabla w, n> + \frac{f\delta\alpha\Delta w}{2l}. \tag{4.2}$$

Thus this estimate of the radial slope of the wavefront suffers nonlinearities when either the *a priori* width of the boundary signal is underestimated or over estimated. Roddier, et al [10] discuss this nonlinearity when the edge is underestimated. For wavefront components with zero Laplacian there is no error by overestimating the width of the edge. However, nonlinearities do appear for signals with nonzero Laplacians. This was observed in [1 O] when a small nonlinearity appeared for defocus even though the sensor was not in saturation. The explanation for this follows from (4.2) where the Lapplacian term is nonzero for defocus, but vanishes for tilt and astigmatism.

The noise characteristics of the radial slope estimate obtained in this manner are addressed in [7]. In summary it is shown that the method has to trade the resolution by which it can ascertain the boundary

displacement with read and dark current noise, and ultimately these noise characteristics are inferior to those of a Hartmann sensor.

5. Wavefront Reconstruction. The problem of wavefront reconstruction is to estimate the wavefront across" an aperture from sampled values. The sampled values are obtained from measurement devices such as a Hartmann sensor, a shearing interferometer, or the curvature sensor as discussed in the previous section. It is important to note that these sensors do not provide direct information of the wavefront, but only first or second derivative information through either the slope or curvature measurement.

The general setup of the reconstruction problem is fairly simple. Let an aperture be defined by a region D in a plane with boundary ∂D. We will be taking D to be a square in most of the analysis to obtain some fairly specific results, but for now we will let D be quite general. Let $w(x), x \in D$ denote the wavefront surface. The reconstruction problem from slope measurements is to determine w(z), $x \in D$ given a sample of the gradient of $w, \nabla w(x_i), i = 1, \ldots .$ n. The problem for curvature sensing is to estimate w(z) given the samples $\Delta w(x_i), i = 1, \ldots .$ n. For curvature sensing it is also necessary to have normal gradient information on the boundary to perform wavefront reconstruction.

For both of these reconstruction problems any constant function can be added to a solution and still solve the system, For the purpose of error analysis it will be shown that the normalizing condition on the reconstructed wave front, $\hat{w}$,

$$\sum_{i,j} \hat{w}_{ij} = 0 \tag{5.1}$$

leads to the minimum variance solution. This solution is also recognized as the "zero piston" solution, and has further relevance to the wavefront correction problem [7].

For more detailed anal ysis, the reconstruction problem will be restricted to a square aperture with regular measurements. We will assume that the square is d x d units and there arc $(N + 1)2$ regularly spaced nodes. We let h denote the mesh width, so that $h = d/N$. We begin with an analysis of slope measurements in this configuration. At each mesh point indexed (i, $j) \in D$ we assume the noisy vector slope measurement

$$[s^x_{ij}\ s^y_{ij}] = \nabla w(x_i,\ y_j) + \eta_{ij} \tag{5.2}$$

where

$$\eta_{ij} = [\eta^x_{ij}\ \eta^y_{ij}]\ , \tag{5.3}$$

with η_{ij} zero mean for every i and j, and with constant covariance $E(\eta^T_{ij}\eta_{ij}) = \sigma^2 I_{2\times 2}$. (The independence of the x and y slope measurements are discussed in [5].)

The gradient $\nabla w(x_i, y_j)$ is approximated at mesh points by the difference operator A_h,

$$(A_h u)(ij) = \left(\frac{u_{ij+1} - u_{ij}}{h}\ \frac{u_{i+1j} - u_{ij}}{h}\right) \tag{5.4}$$

To develop the minimum variance estimator we write the difference operator above as

$$A_h = \frac{1}{h}\begin{bmatrix} A^x \\ A^y \end{bmatrix} \tag{5.5}$$

where

$$(A^x u)(ij) = u_{ij+1} - u_{ij},$$

and

$$(A^y u)(ij) = u_{i+1j} - u_{ij}.$$

Note that

$$A' : \oplus_{i=1,N+1} R^{N+1} \to \oplus_{i=1,N+1} R^N$$

has the (rectangular) block diagonal representation $AZ = \text{diag}(A^x_i)$ with $A^x_i : R^{N+1} \to RN$,

$$A^x_i = \begin{bmatrix} -1 & 1 & 0 & \ldots & 0 \\ 0 & -1 & 1 & \ldots & 0 \\ \vdots & \vdots & \ddots & \ddots & \\ 0 & 0 & \ldots & -1 & 1 \end{bmatrix}.$$

Thus $A'^T: \oplus_{i=1,N+1} R^N \rightarrow \oplus_{i=1,N+1} R^{N+1}$ is also block diagonal with blocks A_i^{xT}. Hence $A^{xT}A^x$ is block, diagonal, with tridiagonal blocks $A_i^{xT}A_i^x$,

$$A_i^{xT}A_i^x = \begin{bmatrix} 1 & -1 & & & \\ -1 & 2 & -1 & & \\ & \ddots & \ddots & \ddots & \\ 0 & \dots & -1 & 2 & -1 \\ 0 & \dots & 0 & -1 & \end{bmatrix} \quad 1$$

A similar, development follows for computing $(A^y)^T A^y$. Here we note that

$$A^y = \begin{bmatrix} -I & I & 0 & \dots & 0 \\ 0 & -I & I & \dots & 0 \\ \vdots & \vdots & \ddots & \ddots & \\ 0 & 0 & \dots & -I & I \end{bmatrix}$$

where I denotes the $(N+1) \times (N+1)$ identity matrix. It follows then that

$$A^{yT}A^y = \begin{bmatrix} 1 & -1 & & & \\ -I & 21 & -I & & \\ & \ddots & \ddots & \ddots & \\ 0 & \dots & -I & 21 & -1 \\ 0 & \dots & 0 & -I & I \end{bmatrix}$$

Since $A^TA = A^{xT}A + A^{yT}A$, we find that A^TA is block tridiagonal

$$A^TA = \begin{bmatrix} B_0 & -I & & & \\ -I & B & -I & & \\ & \ddots & \ddots & \ddots & \\ 0 & \dots & -I & B & -I \\ I & 0 & \dots & 0 & -I & B_0 \end{bmatrix},$$

where B_0 has the tridiagonal form

$$B_0 = \begin{bmatrix} 2 & -1 & & & \\ -1 & 3 & -1 & & \\ & \ddots & \ddots & \ddots & \\ 0 & \dots & -1 & 3 & -1 \\ [0 & \dots & 0 & -1 & 2 \end{bmatrix},$$

and $B = I + B_0$.

The reconstruction problem for slope sensing can be posed as the following minimum variance problem:

$$\min_{\hat{w}} E(|\hat{w} - w|^2) \qquad (5.6a)$$

subject to the constraints

$$E(\hat{w}) = \mathrm{w}, \quad \hat{w} = Ky, \; y = A_h w + \eta, \qquad (5.6b)$$

with $E(\eta) = \mathrm{O}$, and $E(\eta^T\eta) = \mathrm{Q}$, where A_h is defined in (5.5), y is the measurement vector in (5.2), and η is the noise vector from (5.3). If A has full rank the solution to this problem yields the best linear unbiased estimator (BLUE) [25] from the measurements y, and is given as

$$\hat{w} = (A_h^T Q^{-1} A_h)^{-1} A_h^T Q^{-1} y, \qquad (5.7a)$$

with resulting covariance matrix

$$E[(\hat{w} - w)(\hat{w} - w)^T] = (A_h^T Q^{-1} A_h)^{-1}. \tag{5.7b}$$

However, it is evident that Ahasa one dimensional null space spanned by the single vector $v = [1 \ldots 1]^T$ corresponding to a piston mode. We could circumvent this difficult by formulating an equivalent (weighted) least squares problem and derive the estimate via a pseudoinverse solution. But for deriving variance estimates the following approach will be more illuminating. Let $\Psi : R^{(N+1)-1} \to R^{(N+1)}$ be any orthogonal martix $(\Psi^T \Psi = I)$ such that the range of Ψ is the orthogonal complement, call it U, of the subspace spanned by v. We now repose the problem as finding the linear minumum variance estimate in the subspace U:

$$\min_{\hat{w}} \ E(|\hat{w} - w|^2); y = A_h w + \eta \ \text{ where } \ w \in U. \tag{5.8}$$

'l'his is equivalent to the problem "

$$\min_{\hat{u}} E(|\Psi(\hat{u} - u)|^2); \quad y = A_h \Psi u + \eta \tag{5.9}$$

'l'his reposed problem has the interpretation that we are seeking solutions to the problem where the "mean" wavefront is zero, i.e. has zero piston. We will see a little later that the choice Ψ leads in a certain sense to the minimum variance *wavefront* error solution.

The solution to (5.9) above is given by

$$\hat{w} = \Psi \hat{u}, \quad \text{where } \ \hat{u} = [\Psi^T A_h^T Q^{-1} A_h \Psi]^{-1} \Psi^T A_h Q^{-1} y, \tag{5.10}$$

since $\Psi^T A_h^T Q^{-1} A_h \Psi$ is now invertible. The variance of the estimate is

$$E(|w - \hat{w}|^2) = tr\{\Psi[\Psi^T A_h^T Q^{-1} A_h \Psi]^{-1} \Psi^T\} = tr\{\Psi^T \Psi [\Psi^T A_h^T Q^{-1} A_h \Psi]^{-1}\} = tr\{[\Psi^T A_h^T Q^{-1} A_h \Psi]^{-1}\}. \tag{5.11}$$

Observe that since $R(\Psi) = N(A)^{\perp}$, $tr\{[\Psi^T A^T Q^{-1} A \Psi]^{-1}\}$ is the sum of the reciprocals of the nonzero eigenvalues of $A^T Q^{-1} A$. These results arc formally stated in the theorem below.

Theorem 5.1. Let Q denote the variance of the noise term in (5.8), Q = $E(\eta \eta^T)$, and let Ah be the difference operator defined in (5.4). 'l'hen the minimum variance solution is given in (5.10) above, with variance

$$E(|w - \hat{w}|^2) = tr\{[\Psi^t A_h^T Q^{-1} A_h \Psi]^{-1}\}$$

If Q is a scalar matrix, i.e., Q = $\sigma^2 I$ then the variance can be expressed as

$$\begin{aligned} E(|u - \hat{u}|^2) &= \sigma^2 h^2 tr\{[\Psi^t A^T A \Psi]^{-} 1, \\ &= \sigma^2 h^2 \sum \frac{1}{\lambda_i}, \end{aligned}$$

where the λ_i's are the nonzero eigenvalues of $A^T A$.///

We will next see how a very analogous situation develops for curvature sensing when using a 5-point scheme discretization of the Laplacian [15]. A point to keep in mind while we develop the result below is that it is tied to this particular approximation of the Laplacian, and other options for discretization are available for curvature sensing.

Curvature sensing produces the following sampled Laplacian and radial derivative signal:

$$\Delta w_{ij} + \eta_{ij}^{int} = y_{ij}; \qquad \nabla w_{ij} \cdot \text{n+-} \ \eta_{ij}^{slope} = y_{ij}^{slope}$$

Recall here that the noise η has two components; an interior component, η^{int} associated with the curvature signal, and a boundary component η^{slope} associated with the measurement of the normal derivative of the

wave front at the boundary. Discretization via the 5-point scheme leads to the difference equation in the interior of D,

$$\frac{4w_{ij} - w_{ij-1} - w_{ij+1} - w_{i-1j} - w_{i+1j}}{h^2} = \text{Yij} + \eta_{ij}, \tag{5.12}$$

at interior points not adjacent to a boundary, and at points adjacent to the boundary either:

$$\frac{2w_{ij} - w_{ij+1} - w_{i+1\,j}}{h^2} = y_{ij} + \eta_{ij}, \tag{5.13}$$

for corner points, and for points not at the corner:

$$\frac{3w_{ij} - w_{ij-1} - w_{ij+1} - w_{i+1j}}{h^2} = y_{ij} + \eta_{ij}. \tag{5.14}$$

In the adjacent cases the signal is a combination of the curvature signal and slope signal.' For example, for points adjacent to the "north" boundary ($i = 2$)

$$y^{slope} = \frac{w_{i-1j} - w_{ij}}{h} + \eta^{slope}.$$

Hence,

$$w_{i-1j} = w_{ij} + h[y^{slope} - \eta^{slope}],$$

and consequentl y the right sides of (5.13- 5.14) contain the terms y^{slope}/h and η^{slope}/h.

Observe now that this discretized system has the form

$$A_h^T A_h w + \eta = y \tag{5.15}$$

with A_h defined as in (5.5). After introducing the normalizing matrix Ψ for the estimation problem (5.15) above, we obtain the curvature sensing reconstruction analogue to Theorem 5.1.

Theorem 5.2. Let Q denote the variance of the noise in (5.12-5.14) and let A_h be the difference operator defined in (5.5). Then the minimum variance solution to (5.15) is

$$\hat{w} = \Psi[\Psi^T A_h^T A_h \Psi]^{-1}[\Psi^T Q^{-1} \Psi]^{-1} \Psi^T Q^{-1} y,$$

with variance

$$E(|w - \hat{w}|^2) = tr\{[\Psi^T A_h^T A_h Q^{-1} A_h^T A_h \Psi]^{-1}\}.$$

If Q is a scalar matrix, i.e., $Q = \sigma^2 I$ then

$$\hat{w} = \Psi(\psi^T A_h^T A_h \Psi)^{-1} \Psi^T y,$$

and the variance can be expressed as

$$E(|u - \hat{u}|^2) = \sigma^2 h^4 tr\{[\Psi^T A^T A A^T A \Psi]^{-1}\}$$
$$= \sigma^2 h^4 \sum \frac{1}{\lambda_i^2},$$

where the λ_i's are the nonzero eigenvalues of $A^T A$.///

There arc two distinctions between the curvature and slope sensing variance estimates. The first is the factor of h^4 that appears in the curvature sensor reconstruction error, versus the factor of h^2 in the slope sensing reconstruction error. The second difference is that the reconstruction in curvature reconstruction involves the sum over the square of the reciprocals of the eigenvalues of $A^T A$, as opposed to the reciprocals of the eigenvalues for slope sensing. Thus we. see immediatel y that the trade between curvature and slope

sensing is governed by both the growth of the reciprocals of the eigenvalues of the Laplacian and the mesh size. On square domains these eigenvalues can be computed analytically;

Proposition 5.3. The eigenvalues of A^TA are given by

$$\lambda_{ij} = 4 - 2\cos\frac{\pi i}{N+1} - 2\cos\frac{\pi j}{N+1} \quad i,j = 0,\ldots,N.$$

Proof. Recall that $A^TA = A^{xT}A^x + A^{yT}A^y$ *(see* development following (5.5)), Consider the matrix

$$K = \begin{bmatrix} 1 & -1 & & & \\ -1 & 2 & -1 & & \\ & \ddots & \ddots & \ddots & \\ 0 & \ldots & -1 & 2 & -1 \\ 0 & \ldots & 0 & -1 & 1 \end{bmatrix},$$

and let $\{\lambda_i\}_i$, {ui }$_i$ denote the sets of eigenvalues and eigenvectors of this matrix, respectively. Denote the components of the vector u_i by superscript, i.e. $u_i = [u_i^1, \ldots . u_i^n]^T$ where n = N + 1. Now construct the $n^2 x 1$ vector q_{ik},

$$q_{ik} = \begin{bmatrix} u_i^1 u_k \\ u_i^2 u_k \\ \vdots \\ u_i^n u_k \end{bmatrix}.$$

Observe that

$$A^{xT}A^x q_{ik} = \begin{bmatrix} u_i^1 K u_k \\ \vdots \\ u_i^n K u_k \end{bmatrix}. \tag{5.16}$$

Hence, since $Ku_k = \lambda_k u_k$, it follows that

$$A^{xT}A^x = \lambda_k q_{ik}.$$

Observe next that

$$A^{yT}A^y q_{ik} = \begin{bmatrix} u_i^1 v - u_i^2 v \\ \vdots \\ -u_i^{j-1} v + 22u_i^j v - u_i^j u_i^j v. \\ \vdots \\ -u_i^{n-1} v + u_i^n v \end{bmatrix}$$

But since $Ku_i = \lambda_i u_i$, it follows that

$$A^{yT}A^y q_{ik} = \lambda_i q_{ik}. \tag{5.17}$$

From (5.16) and (5.17) we obtain $A^TA = (\lambda_i + \lambda_k)w_{ik}$.

Now K is recognized as tbe stiffness matrix obtained from n unit masses serially coupled by unit stiffness springs unsupported at the ends. The eigenvalues of *K* are given by [1]

$$\lambda_i = 2 - 2\cos\frac{\pi i}{N+1}, i = 0, \ldots, N.$$

It is straightforward to verify that the vectors $\{q_{ik}\}$ are pairwise orthogonal since the set {u$_i$} is an orthornormal set by virtue of the symmetry of *K.* Hence $\{q_{ik}\}$ constitute a complete set of eigenvectors for A^TA, and $\{\lambda_i + \lambda_k\}$ form a complete set of eigenvalues of A^TA.///

We now return to the discussion of the role of the matrix Ψ in the minimum variance solution. There arc of course other ways of, normalizing the problem to obtain unique solutions. For example we could find all solutions in the subspace of solutions that grounds a particular node (i.e., we require $u_{i_0 j_0} = 0$. for some

index $i_0 j_0$.) 'l'his corresponds to selecting another matrix, say $\Gamma : R^{(N+1)^2-1} \to R^{(N+1)^2}$, **such that** $A\Gamma$ **has** full rank. The pertinent observation regarding *any* normalizing matrix Γ such that $\Gamma\bar{v} \neq 0$ is that the associated error variance $tr\{\Gamma[\Gamma^T A_h^T Q^{-1} A_h \Gamma]^{-1}\Gamma^T\}$ has the property that

$$tr\{\Gamma[\Gamma^T A_h^T Q^{-1} A_h \Gamma]^{-1}\Gamma^T\} \geq tr\{[\Psi^T A_h^T Q^{-1} A_h \Psi]^{-1}\}, \tag{5.18}$$

i.e., the variance obtained from the normalization Ψ is a minimum. (The analogous result holds for curvature sensing reconstruction as well.) This result can be deduced from the minimax property [3] of the eigenvalues of a symmetric matrix in the following way: Let $\Gamma = OR$ be a QR factorization of I', i.e. O is an orthogonal matrix and R *is* an invertible triangular matrix. Then

$$\begin{aligned} tr\{\Gamma[\Gamma^T A_h^T Q^{-1} A_h \Gamma]^{-1}\Gamma^T\} &= tr\{OR[R^T O^T A_h^T Q^{-1} A_h OR]^{-1} R^T O^T\} \\ &= tr\{O[O^T A_h^T Q^{-1} A_h O]^{-1} O^T\} \\ &= tr\{[O^T A_h^T Q^{-1} A_h O]^{-1}\}. \end{aligned} \tag{5.19}$$

Writing $\hat{w}_\Gamma$ to denote the estimate 'from the normalization by Γ, we see from (5.19) above that the variance of this estimate is equal to the sum of the reciprocals of the eigenvalues of $O^T A_h^T Q^{-1} A_h O$. *Now* **the** eigenvalues of $\Psi^T A^T Q^{-1} A \Psi$ are **precisely the** nonzero eigenvalues of $A^T Q^{-1} A$, say $\{\lambda_2, \ldots . \lambda_N\}$. By the minimax property, the eigenvalues of $O^T A_h^T Q^{-1} A_h O$, $\lambda'_1, \ldots \lambda'_{N-1}$ satisfy $\lambda'_k \leq \lambda_{k+1}, k = 1, \ldots N-1$. Thus (5.18) follows.

The variances obtained by different normalizations actually have a simple relationship. Recall that the matrix Ψ ensures that the associated estimate $\hat{w}_\Psi$ has the "zero piston" property

$$< \hat{w}_\Psi, v >= 0, \quad \text{where } v = [1 \ldots 1]^T,$$

i.e.,

$$\sum_{ij} \hat{w}_{ij} = 0,$$

where $\hat{w}_{ij}$ denotes the components of $\hat{w}_\Psi$. Although (5.18) holds for any normalization, it can be **shown that**

$$E\{|\hat{w}_\Gamma - w_\Gamma|^2 - (< \bar{v}, \hat{w}_\Gamma - w_\Gamma >)^2\} = E(|\hat{w}_\Psi - w_\Psi|^2),$$

where $\hat{w}_\Psi$ denotes the estimate from using the normalization Ψ [7], What this result says then is that the variance from any normalization of the problem is equal to the variance of any other normalization *after* the error due to the piston component is removed.

6. Comparison of Curvature Sensing and **Hartmann** Sensing Reconstruction Error. Theorems 5.1 and 5.2 together with Proposition 5.3 allow us to make. comparisons between reconstruction error for curvature and slope sensing methods. (For slope sensing we will assume a Hartmann array.) In the case of a square aperture equipped with an $N \times N$ array of subapertures, the covariance matrix Σ for either curvature or Hartmann sensing is related to the eigenvalues $\{\lambda_{kl}\}_{kl}$ of $A^T A$ *as* given in Proposition 5.3. Taking for example, $N = 10$ and $h = 1$ m, for the Hartmann sensor we calculate

$$\begin{aligned} tr(\Sigma_{slope}) &= \sigma^2_{slope} \sum_{k,l} \frac{1}{\lambda_{kl}} \\ &= 67.3\sigma^2_{slope}, \end{aligned}$$

and for the curvature sensor

$$\begin{aligned} tr(\Sigma_{curve}) &= \sigma^2_{curve} \sum_{k,l} \frac{1}{\lambda^2_{kl}} \\ &= 274.5\sigma^2_{curve}, \end{aligned}$$

where σ^2_{slope} denotes the variance of the Hartmann sensor measurement and σ^2_{curve} denotes the variance of the curvature sensor measurement. For the Hartmann array where the centroiding is done. by a quadcell sensor,

$$\sigma_{slope} \approx \frac{3\pi\lambda I_{seeing}}{16\sqrt{\nu}},$$

where λ is the wavelength, I_{seeing} is a variance correction factor given by the Yura approximation for short exposure spot size diameter [9], [19],

$$I_{seeing} = \sqrt{1 + (h/r_0)^2[1 - .37(r_0/h)^{1/3}]}, \quad r_0 = \text{coherence length, h=subaperture size,}$$

and ν is the number of photons per subaperture. Using the values $\lambda = .8 \times 10^{-6}$ m, $h = 1$ m, and $r_0 = .3$ m we get

$$RMS_{slope} = \frac{1.23 \times 10^{-6}}{\sqrt{\nu}}.$$

And using the previous values, j = 150m, and 1 = .05m from Section 3 to characterize σ_{curve}, we obtain the rms estimation error with curvature sensor

$$RMS_{curve} = \frac{3.68 \times 10^{-6}}{\sqrt{\nu}}.$$

Hence, the rms estimation error for curvature sensing for this array is approximately three times larger than for the Hartmann sensor.

We compared reconstruction errors for various array sizes with N = 5, 10, 20, and 40 corresponding to mesh widths of h = 2 m, 1 m, .5 m, and .25 m, respectively. These results are contained in the table below.

Table 6.1. Curvature vs. Slope Sensing Reconstruction

h	.25	.5	1.0	2.0
Hartmann	5.84×10^{-7}	7.87×10^{-7}	1.23×10^{-6}	2.13x 10-6
Curvature	9.07 x 10-'	1.82×10^{-6}	3.68×10^{-6}	7.60×10^{-6}

These results are contrary to what has been reported in the literature regarding the accuracy of curvature sensing and reconstruction ([1 O], [13], [14]). What is observed in the table above is that the curvature sensor compares more favorably with the Hartmann sensor as the resolution increases. The trade that occurs between the two sensing methods is that although reconstruction error from the Laplacian measurement grows more rapidly than the reconstruction error from gradient measurements, this effect is mitigated by the property that the Laplacian is a higher accuracy measurement ($O(h^2)$ versus $O(h)$.) These trades are made more clear below.

Asymptotic estimates for the estimation error can be developed using the characterization of the eigenvalues in Proposition 5.3. For large N we have the approximation

$$4 - 2\cos\frac{\pi k}{N+1} - 2\cos\frac{\pi l}{N+1} \approx \frac{\pi^2(k^2+l^2)}{(N+1)^2}.$$

Hence,

$$\sum_{l,k} \frac{1}{\lambda_{kl}} \approx (N+1)^2 \sum_{k,l=1}^{N} \frac{1}{\pi^2(k^2+l^2)},$$

and

$$\sum_{l,k} \frac{1}{\lambda_{kl}^2} \approx (N+1)^4 \sim \frac{\pi}{4}(k^2+l^2)^{-2}$$

These sums can be approximated as

$$\sum_{l,k} \frac{1}{\lambda_{kl}} \approx O(N^3 \log(N)),$$

and

$$\sum_{l,k} \frac{1}{\lambda_{kl}^2} \approx O(N^4).$$

From these approximations we obtain asymptotic estimates of the surface reconstruction error for slope sensing and curvature sensing:

Theorem 6.1. Let d = length of a square aperture, and let h denote the mesh size. Then for large d/h, the following asymptotic reconstruction error estimates are obtained for slope sensing and curvature sensing, respectively:

$$RMS_{slope} \sigma_{slope} \sqrt{\log(d/h)},$$

and

$$RMS_{curve} \approx h \sigma_{curve} d.$$

Previous analysis of the effect of reconstruction on slope vs. curvature sensing focuses on the noise propagation properties, and essentially fixes the mesh size h while increasing the aperture sized [10], [13], [14]. As d increases (with h fixed) it is seen that the RMS_{slope} grows logarithmically and RMS_{curve} grows linearly as reported in these references. The error propagates differently, however, if wc fix the aperture size and decrease the mesh size. Since for $h < r_0$

$$\sigma_{slope} \approx \frac{3\pi\lambda}{16h} \frac{1}{\sqrt{\nu}},$$

and

$$\sigma_{curve} \approx \frac{l}{f^2} \frac{1}{\sqrt{\nu}},$$

where v again denotes the number of collected photons, wc find that

$$RMS_{slope} = \frac{3\pi\lambda}{16} \sqrt{\log(d/h)} \frac{1}{\sqrt{\nu}},$$

and

$$RMS_{curve} = \frac{lhd}{f^2\sqrt{\nu}}.$$

Thus we see that curvature sensing may actually be superior to Harmann sensing when the subaperture diameters must be small. (Such a circumstance is envisioned for the dense segmented primary mirror of the SELENE telescope [18].)

Acknowledgements

This work was performed at the Jet Propulsion Laboratory, California Institute of Technology, under a contract with the National Aeronautics and Space Administration.

References

[1] R. D. Blevins, "Formulas for Natural Frequency and Mode Shape", Krieger Publishing Co., Malabar, Fl, 1984,

[2] M. Born and E. Wolf, "Principles of Optics", Pergamon Press, oxford, 1989.

[3] G. H. Golub and C. F. Van Loan, "Matrix Computations", Johns Hopkins University Press, Baltimore, 1989.

[4] J. W. Goodman, "Introduction to Fourier Optics", McGraw-Hill, N. Y., 1968.

[5] R. H. Hudgin, Wavefront reconstruction for compensated imaging, J. Opt. Sot.' Am., 67, 1973, pp. 375-378.

[6] K. Ichikawa, A. Lohmann, and M. Takeda, Phase retrieval based on the irradiance transport equation and the Fourie transform method: experiments, Applie Optics, 27, 1988, pp. 3433-3436.

[7] M. Milman, D. Redding, and L. Needels, Natural Guide Star Adaptive Optics Options for Keck Telescope, JPL D-11809, May, 1994.

[8] B. O'Neill, "Elementary Differential Geometry", Academic Press, N. Y., 1966.

[9] R, R. Parenti and R. J. Sasiela, The application of synthetic-beacon technology to astronomy, Technical Report 966, Massachusetts Institute of Technology, Lincoln Laboratory, March, 1993.

[10] F. Roddier, A new curvature method, Applied Optics, 27, 1988, pp. 1223-1225.

[11] C. Roddier, E. Limburg, N. Readier, F. Roddier, M. Northcott, Interferometric imaging through aberrated optics without a reference source, Annual Report SDI/IST Contract, 1989.

[12] F. Roddier, M. Northcutt, J. E. Graves, A simple low-order adaptive optics system for near infrared applications, Pub. Astr. Sot. Pac. 103, Jan., 1991.

[13] C. Roddier and F. Roddier, Wavefront reconstruction from defocused images and the testing of ground-based optical telsecopes, J. Opt. Sot. Amer. (submitted).

[14] J. B. Shellan, pvt. communication (SELENE Project).

[15] G. Strang, "Introduction to Applied Mathematics", Wellesley-Cambridge Press, Wellesley, MA, 1986.

[16] N. Streibel, Phase imaging by the transport equation and intensity, Opt. Comm., 49, 1984, pp. 6–10.

[17] M. R. Teague, Deterministic phase retrieval: A Green's function solution, J. Opt. Sot. Am., 73, 1983, pp. 1434-1441.

[18] R. Ulich and J. D. G. Rather, Innovative approach to next generation telescope design, SPIE Conf. 1236, Tuscon, AZ, 1990.

[19] H. T. Yura, Short-term average optical-beam spread in a turbulent medium, J. Opt. Sot. Am., 63, 1973, pp. 567-572.